ESSAYS AND SCHOLARSHIP

First Things First: Online Advertising Practices and Their Effects on Platform Speech

Crafting policy to combat social media's harmful business practices that encourage destructive speech

BY JEFF GARY & ASHKAN SOLTANI
AUGUST 21, 2019

FREE SPEECH FUTURES
An essay series reimagining the First Amendment in the digital age

Introduction

Problematic and divisive content dominates today's online platforms. Disinformation, abuse, conspiracy theories, and hate speech have risen to prominence and created challenges both online and off. Policymakers, advocates, journalists, and academics have all called on platforms to ensure that this content doesn't overtake more beneficial uses of their services. Many proposed solutions focus on platforms' ability to directly allow or prohibit particular types of speech or users and ask the platforms to do more to remove specific instances of abusive speech.

These proposals raise serious questions about the role a private corporation should play in determining the limits of acceptable speech. Because private companies—such as social media platforms—are not subject to First Amendment restrictions, they alone are able to determine what content is permissible in the public sphere with little restriction, and that dynamic may potentially erode or obviate many of the protections granted under the First Amendment. On the other hand, government regulation or oversight of moderation raises significant questions about whether such restrictions would be permissible under the First Amendment as it is currently understood.

It is our intuition that, rather than focus on particular speakers or speech, policymakers should instead address the underlying platform business practices that encourage much of the most harmful speech. Any long-term solution must address the ways in which platforms collect data and curate content in order to maximally harvest users' attention and sell advertisements, the main method of monetization online.

Specifically, online platforms curate and promote content to drive user engagement (the intensity and frequency with which a user interacts with a platform), which in turn enables platforms to collect more detailed information about the types of content users engage with. This data collection provides platforms with insights not just into the type of content that provokes users to engage *generally*, but also into the preferences, interests, and likely behaviors of *specific users*—insights that then can be used to personalize and deliver the content that is most engaging to each user. Since by definition, the more provocative the content, the more likely it is to garner and retain attention,[1] platforms are economically incentivized to permit and even to encourage the spread of extreme or controversial harmful speech, as it is likely to directly benefit them financially. In a targeted advertising model, misinformation and conspiracy theories are often the product, not an accident.

Addressing targeted advertising as a business practice, rather than specific types of speech and speakers, avoids many of the thorniest First Amendment issues.[2] Removing or lessening platforms' financial incentives to promote bad content to their users dampens questions about the prudence of a private company acting as a de facto censor on public speech and removes concerns over the limits of government commandeering of platforms. At the same time, placing meaningful restrictions on targeted advertising practices addresses many of the structural supports for the wide reach of fake news and harassment.

This paper will explore a framework for regulating platforms without mandating content-based restrictions. To do so, it will explain that the platforms have the technical means to provide curated spaces. The analysis will then show why an ex post content-based moderation system cannot solve the root problems faced by platforms. It will then examine the incentive structure created by the targeted advertising model and show that addressing those incentives would likely reduce the spread of harmful content on many platforms without running afoul of the First Amendment.

Platform Moderation Will Always Be Imperfect

The primary solution proposed to platforms is simply to remove more speech and speakers. Platforms have responded to these proposals and have shown some willingness to ban problematic or fake users outright, build algorithms to detect hate speech and fake news, promote "trusted" news sources and demote others, and employ human moderators.[3] Despite some positive indications, these tools raise significant civil rights and First Amendment concerns on their own—detailed below—and do not always work.[4] Additionally, the platforms have demonstrated that moderation and content removal is often driven by public opinion, rather than established practices, and has not been consistently applied.[5] This section will examine the technical moderation tools available to platforms and analyze their shortcomings.

Many of the current platform moderation approaches can be seen as extensions of early anti-spam techniques developed by email service providers. In the 1990s, email provided a new means to communicate instantly with the world. Naturally, solicitation, fraud, and abuse followed closely behind.[6] Email providers were forced to grapple with unwanted speech and develop technical tools to reduce unwanted messages.[7] As platforms faced similar problems, they turned toward existing techniques to sanitize their services.[8] Platform moderation therefore looks very similar to email spam control: Senders who might have been "blacklisted" on email servers are essentially users banned on platforms. Whitelisted users are now "verified," and whitelisted domains are "trusted sources." Email content filtering has evolved into a complex web of platform policies, filtering techniques, and machine-learning tools.[9]

While both email providers and social media platforms must regulate huge volumes of speech and content, the challenges they face are considerably different, and email-based moderation tools have not risen to the challenge. Unlike email anti-spam techniques, which largely affected private communications, platform moderation plays out in the public eye and directly affects the ability of individuals and entities to speak in public forums, and platform policies shape the degree of protection individual speakers may enjoy.

Blacklisting and Whitelisting

Banning individual users and domains—blacklisting—was one of the first techniques adopted to counter spam.[10] Early email users simply began keeping lists of IP addresses[11] that sent fraudulent emails and prevented those addresses from being able to transfer outgoing emails.[12] This practice, slightly modified, remains a crucial part of platform content moderation: Platforms routinely identify speakers that contravene policies and ban those accounts.[13] Twitter alone removed over 70 million accounts over the course of two months in 2018;[14] Facebook removed 583 million—a quarter of its user base—in the first three months of the same year.

Despite high visibility, this practice has not kept pace with the actual challenges posed by harmful speech on the platforms. For a variety of reasons, platforms have not taken action against many of the most egregious accounts or have only done so after severe public pressure. Many of the accounts identified with promoting "fake news" during the 2016 U.S. presidential election remain online.[16] Banning accounts is also financially disincentivized for platforms, since user growth and engagement are central to their business models: Daily and monthly active users are the primary metrics by which a service's growth is measured.[17] Further, accounts are only banned *after* they have violated a policy—by posting hate speech or advocating violence, for instance. Blacklisting alone is not an effective solution to prevent harmful speech.

Blacklisting also requires platforms to be able to identify users, even pseudonymously, which affects the right to speak anonymously and to associate freely. Additionally, to enable effective blacklisting, platforms must have a record of individual speakers' account

information—such as their IP addresses—to ensure that the speaker does not simply create new accounts and avoid the ban.[18] Resorting to banning users means that platforms cannot provide an arena for anonymous or pseudonymous speech, cornerstones of the First Amendment rights enjoyed by Americans.[19] Platforms' central role also makes blacklisting particularly disruptive. While an email user blacklisted from certain domains was still able to communicate with other parties, a user blacklisted from a platform has no ability to continue communications on that platform. Blacklisting therefore not only is largely ineffective to protect against harmful speech but also erodes constitutionally protected anonymity of speech at a time when the U.S. government has argued that an individual has no right to anonymity when posting online.[20]

Whitelisting on platforms faces different but complementary challenges. Email whitelisting protects legitimate bulk emailers by placing them on pre-approved sender lists, enabling those senders to avoid being filtered as spam.[21] Modern platforms largely mimicked this technique by creating affirmative signals, such as badges and trust-based rankings. One such program, Facebook's "trusted news source" program, ranked news sources and prioritized stories from "trusted news sources," including major outlets such as the *New York Times* and the *Wall Street Journal*. Stories from trusted sources were placed higher in news feeds, displacing other content, such as "blogs that may be on more of the fringe"[23] or spam. These filters rely on determinations about the speaker,[24] rather than the speech itself skirting many of the main challenges on the platforms. Many mainstream news sources and speakers, for instance, promote content that could be considered hateful or misinformation, and promoting those voices absent determinations on their content risks simply lending legitimacy to fringe ideas without addressing other structural issues on platforms.[25]

Platforms also whitelist users in the form of "verified" badges. Like "trusted news sources," verified badges are based on the identity of an individual, rather than any specific speech. While Twitter views verified badges as a simple indication that a particular account is authentic,[26] users may well believe that the checkmark indicates a level of veracity or trust in a verified account.[27] Even more than trusted news, this approach can have significant error costs and promote harmful and abusive content under the guise of "verification." Shock jock Alex Jones, for instance, was verified on Twitter before he was banned, as was aspiring provocateur Jacob Wohl.[28] Whitelisting also removes any right to anonymity an individual might have had, since platforms generally will not verify pseudonymous or anonymous accounts.

Both blacklisting and whitelisting on platforms create significant limitations on the rights of speakers without meaningfully addressing the promulgation and spread of the most harmful types of speech online. Relying on or enhancing these methods is unlikely to lead to sustainable or desirable solutions. Increasing blacklisting, for instance, will not curb harmful speech, as banning occurs only after bad actions. At the same time, an effective banning policy will require platforms to invade the privacy rights of their users and collect personal information to permanently prevent platform access. The whitelisting focus on

speakers, rather than speech, means that divisive or fringe speakers can still promote their theories, only now with an air of legitimacy granted by recognition from the platforms. At the same time, anonymous speech is disadvantaged, since platforms will not verify anonymous accounts and the platform has inserted itself as a gatekeeper to legitimate speech.

Blacklisting also creates immense potential for abuse, and companies have overstepped boundaries in the past. For instance, the Spamhaus Project, an international email reputation service widely used by businesses to filter email, has been accused of abusing its position by blacklisting companies without due process or based on personal animosity.[29] Though Spamhaus, like major social media platforms, protests that it operates "within the boundaries of the law" and remains "accountable to [its] users,"[30] others have alleged that the company uses its powerful position to force other companies to refuse to deal with individuals Spamhaus has decided should be blocked.[31] By virtue of its power to make blacklisting decisions, Spamhaus, a private company, exerts enormous judicial pressure and extra-judicial influence over the business operations of other enterprises with little to no accountability. As platforms continue to consolidate power over online speech, there is significant risk that similar abuses might occur.

Direct Content Filtering

Platforms have also adopted content filtering tools and methods from email providers. Web providers have filtered content since nearly the beginning of the web.[32] Early spam filters relied largely on reading the content of an email to look for keywords or to "score" the message in order to determine whether it was legitimate.[33] For both email and platforms, filtering is central to the product.[34] Without it, there would simply be too much information for consumers to absorb.

Despite the necessity of the practice, filtering creates the most direct potential First Amendment issues. As private actors, platforms can legally remove any content they like. With no external checks, First Amendment protections mean very little, since private companies, not the government, are the bottlenecks to speech. But, if the government were to enforce filters and rules about what speech is permitted, moderation actions would run headlong into existing restrictions on policing speech. Either way, directly moderating content poses serious challenges.

The First Amendment protects only against actions taken by the government. In practice, this means that platforms have wide latitude to set their own policies on what content is and is not acceptable. Platforms do not have to guarantee any procedural rights for users before removing content, nor are platforms subject to formal outside scrutiny of their policies or decision making. Recent strides toward transparency and accountability are encouraging but fall short in that they are not meaningful substitutes for safeguards and checks and balances for users.[35] Indeed, many platform users have been unsuccessful when challenging content removal decisions.[36] These harms are exacerbated by platforms' spotty ability to

remove actual hate speech and disinformation.[37] In response, platforms have proposed new methods for accountability, mimicking existing democratic structures, such as Facebook's proposed "Supreme Court."[38] Unless checked, platforms' nearly limitless power to shape and suppress online speech has the potential to completely swallow large portions of First Amendment protections against the government.

Checking platform moderation power is not a simple proposition. Involving the government in censorship or moderation activities runs headlong into the most salient First Amendment issues. Under current law, the government cannot force a private company to withhold or to publish particular content.[39] More recent cases suggest that the government may not be able to directly prohibit "fake news" or other false speech simply because it is untrue.[40] The complexity of involving the government in platform moderation may be enough to dissuade policymakers from pursuing that course of action. However, if the government does become involved, platforms are likely to bow to governmental pressure in the United States as they have done in other countries where governments have insisted on content concessions.[41] To be sure, the U.S. government has already eroded some legislative speech protections for platforms by effectively insisting that platforms take particular moderation actions against certain types of content.[42]

Overall, direct moderation of content and users creates enormous potential for collision with or usurpation of the protections of the First Amendment. At the same time, it is not even clear that these moderation tools are effective at protecting users and removing illicit content.[43] Direct content moderation is likely an insufficient response to address the speech problems faced by platforms and platform users. The next section will examine how regulating advertising practices online could alleviate many of the most serious challenges online and provide a framework for addressing harmful speech moving forward.

Targeted Advertising Encourages Platforms to Prioritize Controversial Content

Direct moderation of content will likely always fail to achieve the stated goals of the platforms because the business models of the platforms themselves encourage and reward divisive or controversial content. Today, nearly every social media platform makes money selling advertisements, rather than charging individual users.[44] Ad-based platforms auction off users' attention to advertisers and provide the most ad content to users when users spend the most time on the platform.[45] Platforms are therefore incentivized to serve the content that is most engaging and likely to provoke a response, whether that be a share, a like, or a purchase.[46] Time and again, the best fodder for such a response has proved to be incendiary, controversial, and divisive material.[47] For instance, users may share "fake news" even though they know it is factually incorrect when that content reaffirms a user's sense of identity or culture.[48] Platforms themselves may also share or promote controversial content simply to spark discussion or debate, retain user focus, or create engagement.[49] This

discrepancy was made clearer when Facebook changed its content sorting to promote direct content from a user's individual friends over other content. After platform engagement dropped significantly,[50] the company reversed the move. Platforms therefore have significant interest in promoting harmful content in order to keep users engaged as long as possible.

Once a user is engaged, platforms amass enormous amounts of information to target advertisements directly toward what a platform knows a user will respond to. Platforms collect information including social interactions, demographics, what a user clicks and doesn't click, and myriad other data points to build a digital dossier on each user.[51] They know about users' personality types, behavioral quirks, and even emotional states.[52] Platforms know exactly what pushes a user's buttons and how to engage that individual.[53] Platforms then monetize the users by allowing advertisers to engage directly with specific populations. Recent controversies have highlighted the mismatch between business practices and filtering principles: Until recently, Facebook permitted advertisers to use the category "jew haters" as a target for advertisements. It also distributed anti-vaccination advertisements to potential young mothers.[55] Advertisements in this context are not simply a means to sell a product but act to directly shape user behavior off-platform.[56] Unscrupulous advertising practices and content delivery don't just create a harmful platform: They also contribute to off-platform behavior, such as Pizzagate and racial discrimination,[57] and provide harmful advertisers an enormous platform to reach vulnerable populations.[58]

While platforms have paid lip service to reforming content moderation, they have largely avoided making any such commitments with regard to advertising practices.[59] Instead, platforms have doubled down on growth and profits and insisted that additional accountability measures, such as transparency and changes to internal governance, are sufficient to redress content issues online.[60] This is simply not the case. So long as platform profits are reliant on keeping users on-platform as long as possible, controversial and harmful speech will continue to proliferate.

Despite the ubiquity of the targeted advertising, there is growing skepticism that the practice creates more value for anyone outside of the companies providing the targeting. Researchers have found that "behaviourally targeted advertising had increased the publisher's revenue but only marginally" while costing "orders of magnitude" more to deliver.[61] Major platforms have misled brands and advertisers about the value of placing ads.[62] Recently, several large brands have scaled back or stopped the practice altogether. Procter & Gamble reduced its targeted-ad budget in 2018 after it determined that the program was a waste of money.[63] Following the enactment of the European General Data Protection Regulation (GDPR), Google is supporting non-targeted ads in Europe,[64] and the *New York Times* stopped behavioral advertising altogether while raising its overall advertising revenue in the same period.[65] The past several years have seen a reckoning with targeted advertising from within the industry, and it is past time that conversation took hold in the public.

Regulating Advertising Practices Reduces Constitutional Friction

This essay began by outlining the reasons that speech- and user-focused platform moderation are unlikely to be successful long-term solutions and why the practices create maximal challenges to the First Amendment as it is currently understood. It then discussed why platform business practices create incentives and opportunities for platforms to accept and promote controversial content online. The essay will now outline several proposals for addressing the mismatch between platform and public incentives and explain how each approach might stem the flow of unwanted content online.

Congress should restrict platform practices as part of any privacy legislation.

Following a scad of high-profile platform scandals, Congress is considering a number of proposals to increase privacy and data security for U.S. consumers. Despite mounting evidence that platform structure incentivizes privacy-invasive practices,[66] legislators have largely ignored that relationship in public text. While the relationship between privacy and targeted ads is distinct from questions of platform moderation, the two are closely related. Just as targeted ads incentivize platforms to serve divisive content, so too does the model incentivize platforms to gather as much personal information as possible on users without regard for its provenance or future use.[67] Congress and state legislatures should ensure that privacy laws meaningfully address the advertising models of platforms and place restrictions on how data can be collected and used for advertising purposes.

Existing law provides guidance for how Congress might structure such restrictions. For example, the GDPR places restrictions on what data companies may acquire, and once collected, how those companies may use the data.[68] Congress could address some of the most egregious platform behavior by ensuring that, for instance, certain types of information are not collected by platforms and, when data is collected, that not all data can be used for advertising or commercial purposes.[69] Additionally, restrictions could be placed on using information collected in one context from being used in a different context—for example, prohibiting using phone numbers collected for account recovery from being used for cross-platform advertising purposes.[70] The devil is in the details, of course, but focusing on business behavior rather than consumer harms provides a strong basis for addressing platform harms generally and would lessen the need for aggressive moderation.

The California Consumer Privacy Act (CCPA) also provides a starting point for regulators. While many of the harms from divisive speech begin with platforms promoting divisive content to engage users, others result from sharing consumer data with third parties. The CCPA addresses this type of harm by restricting the types of data that platforms and other online providers can share with third parties and placing affirmative obligations on the third

parties once they have received that information.[71] This should, in turn, reduce the ability for third parties to amass detailed profiles about individuals across multiple sites, services, and devices and significantly constrain the ability to influence those users.

The Federal Trade Commission has authority to address advertising practices.

Congress may or may not decide to approach platform advertising head-on. Until they do, or if they do not, the Federal Trade Commission (FTC) likely has existing authority to curtail the targeted advertising practices. Under Section 5 of the FTC Act,[72] the Commission may enjoin "unfair or deceptive" practices. While claims under the so-called "deceptive authority" are more common,[73] the Commission here may well be able to exercise its authority to ban unfair practices.

In an unfairness case, the FTC must show that (1) the injury to consumers is substantial, (2) any injury is not outweighed by other benefits, and (3) the injury is not reasonably avoidable by consumers.[74] As explained above, targeted advertising has the potential to create substantial injury both to individual consumers and to society.[75] Individual consumers may also find themselves excluded from seeing certain ads based on protected categories.[76] While societal harms are more diffuse, they are evident through effects such as filter bubbles, radicalization,[77] and other off-platform activities.[78] These negative effects, though diffuse, provide ample evidence that platform business activities may not provide much of their promised benefit to consumers. Against direct and diffuse harms, there is a growing consensus that targeted advertising may not provide financial benefit to anyone other than the ad-tech companies.[79] Researchers have shown that behavioral advertising likely only creates nominal value compared to contextual or direct advertising and costs significantly more to produce and deliver.[80] A recent empirical study showed that targeted advertisements are, on average, only worth $0.00008 (4%) more than non-targeted ads to the publishers of the advertisements.[81] The risk of data breach or theft to platforms and users is significant, as websites serving targeted advertisements must create, maintain, and store enormous datasets on millions of users.[82] There is a limit to what type of targeted advertising users will even tolerate,[83] and any purported benefit to consumers is likely outweighed by the harm suffered individually and in the aggregate. Lastly, consumers often have no way to avoid harm other than by not using platforms or playing a cat-and-mouse game with online ad blockers. While this is feasible for some, it is not a realistic option in many cases.[84] Platforms themselves do not even have the capability to reduce the potential for harm once consumer information is in the wild.[85]

Depending on the explicit or implicit representations made by platforms to their users, the Commission may also be able to pursue action based on its deceptive practices authority. For example, if the Commission determined that a platform held itself out as a neutral arbiter of discourse but was instead actively promoting certain viewpoints over others, that discrepancy may rise to the level of a violation of the FTC Act. The Commission has sought

enforcement in other situations when advertisers misrepresented the origins and nature of advertisements to consumers.

While additional facts may be necessary to fully support an enforcement action, the Commission certainly has the tools and authorities to explore whether it can bring an action under the unfairness authority. It should do so.

Platforms may face liability despite the Communications Decency Act.

Section 230 of the Communications Decency Act (CDA 230)[87] protects online service providers from being liable for content posted by their users.[88] Under the Act, providers are not treated as the "publishers" of users' speech and therefore are not responsible for its content.[89] This protection applies even when providers restrict access to material, even if that material is constitutionally protected.[90] This provision has often been held out as a form of absolute immunity for social media platforms and other providers for the information on their platforms.[91]

However, Section 230 may not extend as broadly as some claim. Notably, while CDA 230 protects platforms from liability for content published by third parties content hosted by the platform, those protections may not extend to platforms for decisions about business practices and other non-content activities if the practices are not explicitly a publication activity by the platform or a third party.[92] For instance, platforms may be liable for enabling discrimination or other illegal practices by providing certain ad categories to third parties, since creating and monetizing the categories is not a publication activity.[93] Similarly, legal commenters have noted that Section 230 is accorded a much more sweeping effect than its plain language alone might support.[94] While platforms most likely cannot be held accountable for specific instances of hate speech or harassment,[95] they may well face liability for non-speech business practices that enable and encourage malicious or illegal content.

In addition to current liability for non-publication business practices, we should consider whether platforms should face liability for content when the platform has substantially transformed the presentation beyond the contents' original form or context. Platforms do not simply pass along user content—they take an active role in how, when, and where that content is displayed (including selectively not displaying new content or promoting older content repeatedly). To do this, platforms use information about individuals in combination with aggregate information about group behavior (so-called "big data") as well as inferences made about a particular user's behavior (personalized content). They use that information to create insights about likely responses and engagement to particular material and then curate the universe of user-generated content to maximize engagement independent of the original chronology and intent of the content creator.

The platform editorial process does not simply present users with posts created by their friends or connections: It substantially modifies the order, context, and meaning of user content. Platforms are aware of this transformative effect and have acknowledged that their role is more akin to a "media company" than just a tech platform.[96] Twitter has considered surrounding "fake news" with posts debunking false assertions, an acknowledgment that context and position change meaning. Platforms ought not be held liable for every piece of harmful or malicious content generated by their users, but we should seriously consider whether their content curation functions should fall outside the protective embrace of Section 230.

Conclusion

The proposals outlined above are intended to show that many of the challenges posed by speech on platforms—and platforms' responses— may be partially be addressed by legal and regulatory tools that encroach less immediately on the First Amendment. The authors believe that existing constitutional protections likely remain sufficient to address new types and controllers of speech online. However, before addressing whether speech online must be further curtailed, or grappling with thorny questions of which entities are best able to make content determinations, it is first necessary to challenge structural practices that encourage harmful speech.

To this end, the authors have suggested taking concrete steps to better align private and public interests. By focusing on platform practices, we have suggested that many of the incentives for promoting harmful speech can be removed and platforms' priorities more closely aligned with those of the public. That realignment may serve to reduce or obviate many of the most challenging questions thrust into sharp relief by the proliferation of speech online. These solutions are neither complete nor perfect. But they represent actions necessary to understand the types and degree of harms and challenges that do truly exist and provide a framework to grapple with larger questions about the role of platforms in enabling and policing public speech.

---

1 Jack M. Balkin, Fixing Social Media's Grand Bargain 2 (Hoover Inst., Aegis Paper No. 1814, 2018), https://www.hoover.org/sites/default/files/research/docs/balkin_webreadypdf.pdf; James Grimmelmann, *The Platform Is the Message*, 2 Geo. L. Tech. Rev. 217, 230 (2018), https://georgetownlawtechreview.org/wp-content/uploads/2018/07/2.2-Grimmelmann-pp-217-33.pdf; *see* Jack Nicas, *How YouTube Drives People to the Internet's Darkest Corners*, Wall St. J. (Feb. 7, 2018), https://www.wsj.com/articles/how-youtube-drives-viewers-to-the-internets-darkest-corners-1518020478. Indeed, Facebook CEO Mark Zuckerberg has acknowledged this problem, saying, "One of the biggest issues social networks face is that, when left unchecked, people will engage disproportionately with more sensationalist and provocative content. [. . .] This is a basic incentive problem." Mark Zuckerberg, *A Blueprint for Content Governance and Enforcement*, Facebook (Nov. 15, 2018), https://www.facebook.com/notes/mark-zuckerberg/a-blueprint-for-content-governance-and-enforcement/10156443129621634.

2 For instance, Facebook recently settled claims that it aided housing, employment, and credit discrimination and is facing a separate lawsuit by the U.S. Department of Housing and Urban Development alleging housing discrimination in its advertisement practices. Colin Lecher, *Facebook Drops Targeting Options for Housing, Job, and Credit Ads After Controversy*, The Verge (Mar. 19, 2019), https://www.theverge.com/2019/3/19/18273018/facebook-housing-ads-jobs-discrimination-settlement; Katie Benner, Glenn Thrush, and Mike Isaac, *Facebook Engages in Housing Discrimination with Its Ad Practices, U.S. Says*, N.Y. Times (Mar. 28, 2019), https://www.nytimes.com/2019/03/28/us/politics/facebook-housing-discrimination.html.

3 *See* Mark Zuckerberg, Facebook (Dec. 2, 2018), https://www.facebook.com/4/posts/10105865715850211; Tony Romm & Elizabeth Dwoskin, *Jack Dorsey Says He's Rethinking the Core of How Twitter Works*, Wash. Post (Aug. 15, 2018), https://www.washingtonpost.com/technology/2018/08/15/jack-dorsey-says-hes-rethinking-core-how-twitter-works.

4 Jason Koebler & Joseph Cox, *The Impossible Job: Inside Facebook's Struggle to Moderate Two Billion People*, Motherboard (Aug. 23, 2018), https://motherboard.vice.com/en_us/article/xwk9zd/how-facebook-content-moderation-works.

5 For instance, Facebook's recent removal and reinstatement of Elizabeth Warren's presidential campaign ad highlighted platforms' inconsistent moderation practices and susceptibility to public pressure. *See* Brian Feldman, *That Facebook Accidentally Removed Elizabeth Warren's Ads Is the Point*, N.Y. Mag. (Mar. 12, 2019), https://nymag.com/intelligencer/2019/03/facebook-accidentally-removed-elizabeth-warrens-anti-fb-ads.html. *See also* Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 Harv. L. Rev. 1598 (2018).

6 Finn Brunton, Spam: A Shadow History of the Internet 63–77 (2013).

7 S. Rep. No. 108-102, at 6 (2003) ("ISPs must respond to rising volumes of spam by investing in new equipment to increase capacity and customer service personnel. [. . .] ISPs also face high costs maintaining e-mail filtering systems and other anti-spam technology on their networks to reduce the deluge of spam."), available at https://www.congress.gov/108/crpt/srpt102/CRPT-108srpt102.pdf.

8 Matt Hicks, *Explaining Facebook's Spam Prevention Systems*, Facebook (June 29, 2010), https://www.facebook.com/notes/facebook/explaining-facebooks-spam-prevention-systems/403200567130.

9 Paul Gillin, *The Art and Science of How Spam Filters Work*, SecurityIntelligence (Nov. 2, 2016), https://securityintelligence.com/the-art-and-science-of-how-spam-filters-work.

10 *See* Luke Martinez, *What Are Email Blacklists and How to Avoid Them,* SendGrid (Feb. 2, 2019), https://sendgrid.com/blog/email-blacklist.

11 The unique identifier of a particular machine connected to the internet. Stephanie Crawford, *What Is An IP Address?*, HowStuffWorks (Jan. 12, 2001), https://computer.howstuffworks.com/internet/basics/what-is-an-ip-address.htm.

12 J. Levine, *DNS Blacklists and Whitelists*, Internet Res. Task Force (2010), https://tools.ietf.org/html/rfc5782. While spammers on blacklisted servers can still send email anywhere, the receiving server checks incoming messages to see if they have originated at blacklisted addresses. Logan Harbaugh, *Inside the Spam Filter*, InfoWorld (Nov. 14, 2003), https://www.infoworld.com/article/2678975/networking/inside-the-spam-filter.html.

13 Marissa Lang, *Blocked and Banned by Social Media: When Is It Censorship?*, S.F. Chron. (Aug. 30, 2016), https://www.sfchronicle.com/business/article/Blocked-and-banned-by-social-media-When-is-it-9193998.php.

14 Craig Timberg & Elizabeth Dwoskin, *Twitter Is Sweeping Out Fake Accounts Like Never Before, Putting User Growth at Risk*, Wash. Post (July 6, 2018), https://www.washingtonpost.com/technology/2018/07/06/twitter-is-

sweeping-out-fake-accounts-like-never-before-putting-user-growth-risk.

15 Alfred Ng, *Facebook Deleted 583 Million Fake Accounts in the First Three Months of 2018*, CNet (May 15, 2018),

16 For example, as recently as 2018, over 80% of Twitter accounts that posted "fake news" during the 2016 election remained active. *Disinformation, 'Fake News' and Influence Campaigns on Twitter* (2018), Knight Found., https://kf-site.production.s3.amazonaws.com/media_elements/files/000/000/238/original/KF-DisinformationReport-final2.pdf.

17 *See* Dante Disparte, *Facebook and the Tyranny of Monthly Active Users*, Forbes (July 28, 2018), https://www.forbes.com/sites/dantedisparte/2018/07/28/facebook-and-the-tyranny-of-monthly-active-users/#161764ba6aea ("Chasing user growth at all costs while concurrently developing tenuous business models centered on monetizing user data led Facebook and Twitter to commit some seemingly unforgivable (and unforgettable) sins, at least in the meantime.").

18 Oliver Darcy, *Facebook Removes 22 Pages Linked to Conspiracy Theorist Alex Jones and InfoWars*, CNN (Feb. 5, 2019), https://www.cnn.com/2019/02/05/media/facebook-alex-jones-infowars-pages.

19 *E.g.*, McIntyre v. Ohio Elections Commission, 514 U.S. 334 (1995) (state law banning anonymous political speech was unconstitutional); NAACP v. Alabama, 357 U.S. 449 (1958) (individuals have right to associate anonymously with political organizations). Reducing or removing online anonymity also disproportionately impacts at-risk communities and individuals. *See* Sara Baker, *Why Online Anonymity Is Critical for Women*, Women's Media Ctr. (Mar. 11, 2016), http://www.womensmediacenter.com/speech-project/why-online-anonymity-is-critical-for-women.

20 Kurt Wagner, *Twitter Fights to Protect Anonymous Users More Often Than You'd Think*, Recode (Apr. 10, 2017), https://www.recode.net/2017/4/10/15244754/twitter-lawsuit-government-anonymous-users.

21 Clea Moore, *Email Whitelists: How They Work*, SparkPost (Apr. 24, 2017), https://www.sparkpost.com/blog/how-whitelists-work.

22 Lydia Polgreen, *CEO Mark Zuckerberg Says Facebook Will Rank News Outlets By Trustworthiness*, Huffington Post (May 1, 2018), https://www.huffingtonpost.com/entry/mark-zuckerberg-facebook-media-trust_us_5ae90e25e4b00f70f0ed0725; Aja Romano, *Mark Zuckerberg Lays Out Facebook's 3-Pronged Approach to Fake News*, Vox (Apr. 3, 2018),

23 Romano, *supra* note 22.

24 Adam Rogers, *You Can't Trust Facebook's Search for Trusted News*, Wired (Jan. 25, 2018), https://www.wired.com/story/you-cant-trust-facebooks-search-for-trusted-news.

25 *See, e.g.*, the recent hearings held by the House of Representatives on extremism on social media, where a witness invited by House Republicans repeated conspiracy theories often disseminated online. FOX 10 Phoenix, *WATCH: Candace Owens Opening Statement at U.S. House Hearing*, YouTube (Apr. 9, 2019), https://www.youtube.com/watch?v=RmuFIM4meXg. This statement was picked up by local Fox affiliates, who pushed it further mainstream by publication and reference on social media. Alice E. Marwick, *Why Do People Share Fake News? A Sociotechnical Model of Media Effects*, 2 Geo L. Tech Rev. 474, 494 (2018) ("Understanding Fox News is *extremely* important to understanding problematic information. Fox, along with its pundits and local affiliates, often amplifies far-right stories that begin in fringe online communities.") (emphasis in original).

26 *About Verified Accounts*, Twitter Help Ctr., https://help.twitter.com/en/managing-your-account/about-twitter-verified-accounts ("The blue verified badge on Twitter lets people know that an account of public interest is authentic. [. . .] A verified badge does not imply an endorsement by Twitter.").

27 Research in online shopping, for instance, shows that consumers are more likely to engage when they trust a page and have a "gut feeling" that a page is legitimate, often determined by "how visually secure" the page looks. A verified badge provides a clear—though perhaps unintentional—indicator of visual security to Twitter users and perhaps operates in much the same way. *See* Jamie Appleseed, *How Users Perceive Security During the Checkout Flow (Incl. New "Trust Seal" Study)*, Baymard Inst. (Oct. 5, 2016).

28 *See* Lorraine Longhi, *Pro-Trump Twitter Figure Jacob Wohl Was Accused of Defrauding Arizona Investors*, WUSA9 (Nov. 2, 2018), https://www.wusa9.com/article/news/nation-now/pro-trump-twitter-figure-jacob-wohl-was-accused-of-defrauding-arizona-investors/465-23bc177c-231a-42d0-a510-15138a6a2421; Kelvin Chan, *Twitter CEO Defends Decision Not To Ban Alex Jones, Infowars*, Chi. Trib. (Aug. 8, 2018), https://www.chicagotribune.com/business/ct-biz-twitter-alex-jones-infowars-20180808-story.html.

29 Peter Bright, *When Spammers Go to War: Behind the Spamhaus DDoS*, ArsTechnica (Mar. 28, 2013), https://arstechnica.com/information-technology/2013/03/when-spammers-go-to-war-behind-the-spamhaus-ddos.

30 Lucian Constantin, *DDoS Attack Against Spamhaus Was Reportedly the Largest In History*, ITWorld (Mar. 27, 2013), https://www.itworld.com/article/2714146/ddos-attack-against-spamhaus-was-reportedly-the-largest-in-history.html.

31 *Id.*

32 *DNS Blacklists and Whitelists*, Internet Res. Task Force (Feb. 2010), https://tools.ietf.org/html/rfc5782 (noting that a popular IP blacklist started in 1997 quickly became the basis for IP-based blocking commonly used today).

33 *See* Ben, *Most Common Spam Filter Triggers*, MailChimp (Feb. 4, 2009), https://mailchimp.com/resources/most-common-spam-filter-triggers (i.e., using "Dear" as a salutation increased the likelihood a message would be counted as spam, as did using the phrase "extra inches.").

34 Tarleton Gillespie, *Platforms Are Not Intermediaries*, 2 Geo. L. Tech. Rev. 198 (2018).

35 For instance, Facebook takes active steps to prevent journalists from using automated tools to conduct research about Facebook practices. *See* Jeremy B. Merrill & Ariana Tobin, *Facebook Moves to Block Ad Transparency Tools—Including Ours*, ProPublica (Jan. 28, 2019), https://www.propublica.org/article/facebook-blocks-ad-transparency-tools.

36 Caroline Haskins, *86 Organizations Demand Zuckerberg to Improve Takedown Appeals*, Motherboard (Nov. 15, 2018), https://motherboard.vice.com/en_us/article/vbadzy/86-organizations-demand-zuckerberg-to-improve-takedown-appeals.

37 *See* Yair Rosenberg, *Confessions of a Digital Nazi Hunter*, N.Y. Times (Dec. 27, 2017), https://www.nytimes.com/2017/12/27/opinion/digital-nazi-hunter-trump.html.

38 Kate Klonick & Thomas Kadri, *How to Make Facebook's "Supreme Court" Work*, N.Y. Times (Nov. 17, 2018), https://www.nytimes.com/2018/11/17/opinion/facebook-supreme-court-speech.html.

39 Miami Herald Publishing Co. v. Tornillo, 418 U.S. 241 (1974) (state law requiring newspapers to provide equal coverage to opposing political views was unconstitutional); Near v. Minnesota, 283 U.S. 697 (1931) (state law requiring newspapers to receive government permission to print content was unconstitutional).

40 United States v. Alvarez, 567 U.S. 709 (2012) (holding portions of "False Valor Act" unconstitutional with four justices agreeing that false speech is protected speech, though noting that false statements made for material gain may be viewed differently).

41 *See, e.g.*, Eva Galperin, *Facebook Caves to Turkish Government Censorship*, Elec. Frontier Found. (Jan. 29, 2015), https://www.eff.org/deeplinks/2015/01/facebook-caves-turkish-government-censorship.

42 For instance, the recent SESTA/FOSTA amendment to the Communications Decency Act, 47 U.S.C. § 230, removed immunity for websites suspected of permitting posts from sex traffickers or about trafficked people. Though presumably well intentioned, the move has increased rates of violence against sex workers. Nicole Karlis, *For Bay Area Sex Workers, a New Federal Law Means Less Safety and More Poverty*, Salon (Nov. 5, 2018), https://www.salon.com/2018/11/04/for-bay-area-sex-workers-a-new-federal-law-means-less-safety-and-more-poverty.

43 Cindy Cohn, *Bad Facts Make Bad Law: How Platform Censorship Has Failed so Far and How to Ensure that the Response to Neo-Nazis Doesn't Make it Worse*, 2 Geo. L. Tech. Rev. 432 (2018).

44 Sean Burch,*"Senator, We Run Ads": Hatch Mocked for Basic Facebook Question to Zuckerberg*, SFGate (Apr. 10, 2018), https://www.sfgate.com/entertainment/the-wrap/article/Senator-We-Run-Ads-Hatch-Mocked-for-Basic-12822523.php.

45 Darla Cameron, *How Targeted Advertising Works*, Wash. Post (Aug. 22, 2013), https://www.washingtonpost.com/apps/g/page/business/how-targeted-advertising-works/412; Shoshana Zuboff, *"Surveillance Capitalism" Has Gone Rogue. We Must Curb Its Excesses.*, Wash. Post (Jan. 24, 2019), https://www.washingtonpost.com/amphtml/opinions/surveillance-capitalism-has-gone-rogue-we-must-curb-its-excesses/2019/01/24/be463f48-1ffa-11e9-9145-3f74070bbdb9_story.html.

46 Balkin, *supra* note 1.

47 Dick Lily, *Social Media's Algorithms Lead Us Down Dark, Divisive Rabbit Holes*, Seattle Times (Oct. 28, 2018), https://www.seattletimes.com/opinion/social-medias-algorithms-lead-us-down-dark-divisive-rabbit-holes.

48 Alice E. Marwick, *Why Do People Share Fake News? A Sociotechnical Model of Media Effects*, 2 Geo. L. Tech. Rev. 474, 476-77 (2018).

49 Alexis C. Madrigal, *Why Conspiracy Theory Videos Work So Well on YouTube*, The Atlantic (Feb. 21, 2019), https://www.theatlantic.com/technology/archive/2019/02/reason-conspiracy-videos-work-so-well-youtube/583282.

50 Todd Spangler, *Facebook Stock Slumps After Mark Zuckerberg Signals Major Changes to News Feed*, Variety (Jan. 12, 2018), https://variety.com/2018/digital/news/facebook-stock-mark-zuckerberg-news-feed-1202662782.

51 *Facebook, Inc. Responses to Questions for the Record from U.S. Senate Committee on the Judiciary 84, 141,* Scribd (June 8, 2018), https://www.scribd.com/document/381569036/Zuckerberg-Responses-to-Judiciary-Committee-QFRs#download&from_embed ("We collect the content, communications and other information users provide when they use our Products, including when they sign up for an account, create or share content, and message or communicate with others…[W]e use information collected about a person's use of our Products on their phone to better personalize the content (including ads) [. . .] or to measure whether they took an action in response to an ad we showed them on their phone or on a different device.").

52 *Id.* at 92 ("Facebook does research in a variety of fields. [. . .] If proposed work is focused on studying particular groups or populations (such as people of a certain age) or if it relates to content that may be considered deeply personal (such as emotions) it will go through an enhanced review process before research can begin.").

53 Louise Matsakis, *Facebook's Targeted Ads Are More Complex Than It Lets On*, Wired (Apr. 25, 2018), https://www.wired.com/story/facebooks-targeted-ads-are-more-complex-than-it-lets-on.

54 Julia Angwin et al., *Facebook Enabled Advertisers to Reach "Jew Haters"*, ProPublica (Sept. 14, 2017),

55 Meira Gebel, *Anti-Vaccination Ads on Facebook Are Targeting Pregnant Women, While a Measles Outbreak Spreads Across the Country*, Bus. Insider (Feb. 14, 2019), https://www.businessinsider.com/anti-vaccine-facebook-ads-target-pregnant-women-as-measles-spreads-2019-2.

56 Rebecca Walker Reczek, Christopher Summers & Robert Smith, *Targeted Ads Don't Just Make You More Likely to Buy—They Can Change How You Think About Yourself*, Harv. Bus. Rev. (Apr. 4, 2016), https://hbr.org/2016/04/targeted-ads-dont-just-make-you-more-likely-to-buy-they-can-change-how-you-think-about-yourself.

57 Charles V. Bagli, *Facebook Vowed to End Discriminatory Housing Ads. Suit Says It Didn't.*, N.Y. Times (Mar. 27, 2018), https://www.nytimes.com/2018/03/27/nyregion/facebook-housing-ads-discrimination-lawsuit.html; Natasha Lomas, *The Facts About Facebook*, TechCrunch (Jan. 26, 2019), https://techcrunch.com/2019/01/26/the-facts-about-facebook.

58 Zeke Faux, *How Facebook Helps Shady Advertisers Pollute the Internet*, Bloomberg (Mar. 27, 2018), https://www.bloomberg.com/news/features/2018-03-27/ad-scammers-need-suckers-and-facebook-helps-find-them.

59 In response to recent controversy over anti-vaccination advertisements, Facebook has taken steps to reduce paid anti-vax content. On the other hand, the platform has refused to remove pages and groups dedicated to the subject. Rachel Becker, *Facebook Outlines Plans to Curb Anti-Vax Conspiracy Theories*, The Verge (Mar. 7, 2019), https://www.theverge.com/2019/3/7/18255107/facebook-anti-vaccine-misinformation-measles-outbreaks-group-page-recommendations-removal.

60 *See, e.g.*, *Making Pages More Transparent and Accountable*, Facebook Newsroom (Jan. 23, 2019), https://newsroom.fb.com/news/2019/01/making-pages-more-transparent.

61 Natasha Lomas, *The Case Against Behavioral Advertising Is Stacking Up*, TechCrunch (Jan. 20, 2019), https://techcrunch.com/2019/01/20/dont-be-creepy.

62 Suzanne Vranica & Jack Marshall, *Facebook Overestimated Key Video Metric for Two Years*, Wall St. J. (Sept. 22, 2016), https://www.wsj.com/articles/facebook-overestimated-key-video-metric-for-two-years-1474586951.

63 David Dayen, *Ban Targeted Advertising*, New Republic (Apr. 10, 2018), https://newrepublic.com/article/147887/ban-targeted-advertising-facebook-google.

64 Erica Sweeney, *Google Will Support Non-Targeting Ads for GDPR Compliance*, Marketing Dive (Mar. 26, 2018), https://www.marketingdive.com/news/google-will-support-non-targeting-ads-for-gdpr-compliance/519912.

65 Jessica Davies, *After GDPR, The New York Times Cut Off Ad Exchanges In Europe—And Kept Growing Ad Revenue*, Digiday (Jan. 16, 2019), https://digiday.com/media/new-york-times-gdpr-cut-off-ad-exchanges-europe-ad-revenue.

66 *See* Shoshana Zuboff, The Age of Surveillance Capitalism (2019).

67 Sam Schechner & Mark Secada, *You Give Apps Sensitive Personal Information. Then They Tell Facebook.*, Wall St. J. (Feb. 22, 2019), https://www.wsj.com/articles/you-give-apps-sensitive-personal-information-then-they-tell-facebook-11550851636.

68 2 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119/1) 5, 6 ("Personal

information shall be [. . .] collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes.") (Article 6 lays out the bases for lawful processing of personal information).

69 For instance, greater attention should be paid to the ways in which companies use inferences from collected personal information to further enhance profiles on users, often without explicit consent or awareness of the individuals affected. *See* Michael Kassner, *Unintended Inferences: The Biggest Threat to Data Privacy and Cybersecurity*, Tech Republic (Mar. 10, 2019), https://www.techrepublic.com/article/unintended-inferences-the-biggest-threat-to-data-privacy-and-cybersecurity.

70 Natasha Lomas, *Yes Facebook Is Using Your 2FA Phone Number to Target You With Ads*, TechCrunch (Sept. 27, 2018), https://techcrunch.com/2018/09/27/yes-facebook-is-using-your-2fa-phone-number-to-target-you-with-ads.

71 California Assembly Bill No. 375 (California Consumer Privacy Act) (2018), https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180AB375.

72 15 U.S.C. § 45(a).

73 *See, e.g.*, Cobun Keegan & Calli Schroeder, *Unpacking Unfairness: The FTC's Evolving Measures of Privacy Harms*, 15 J.L. Econ. Pol'y 19 (2019) (noting that "the FTC has been cautious in its use of charges of unfairness in consumer protection enforcement—instead relying primarily on charges of deception.").

74 Fed. Trade Comm'n, Advertising and Marketing on the Internet (2000), https://www.ftc.gov/system/files/documents/plain-language/bus28-advertising-and-marketing-internet-rules-road2018.pdf.

75 *See, e.g.*, Charles V. Bagli, *Facebook Vowed to End Discriminatory Housing Ads. Suit Says It Didn't.*, N.Y. Times (Mar. 27, 2018), https://www.nytimes.com/2018/03/27/nyregion/facebook-housing-ads-discrimination-lawsuit.html.

76 Ariana Tobin, *Facebook Promises to Bar Advertisers From Targeting Ads by Race or Ethnicity. Again.*, ProPublica (July 25, 2018), https://www.propublica.org/article/facebook-promises-to-bar-advertisers-from-targeting-ads-by-race-or-ethnicity-again.

77 Mathew Ingram, *YouTube's Secret Life as an Engine for Right-Wing Radicalization*, Colum. Journalism Rev. (Sept. 19, 2018), https://www.cjr.org/the_media_today/youtube-conspiracy-radicalization.php.

78 *See, e.g.*, Kristine Phillips & Brian Fung, *Facebook Admits Social Media Sometimes Harms Democracy*, Wash. Post (Jan. 22, 2018), https://www.washingtonpost.com/news/the-switch/wp/2018/01/22/facebook-admits-it-sometimes-harms-democracy; Amanda Taub & Max Fisher, *Facebook Fueled Anti-Refugee Attacks in Germany, New Research Suggests*, N.Y. Times (Aug. 21, 2018), https://www.nytimes.com/2018/08/21/world/europe/facebook-refugee-attacks-germany.html.

79 *See* Chris Matthews, *DuckDuckGo: We're Profitable Without Tracking You*, Fortune (Oct. 9, 2015), http://fortune.com/2015/10/09/duckduckgo-profitable.

80 Lomas, *supra* note 61 ("[R]esearch showed that behaviourally targeted advertising had increased the publisher's revenue but only marginally. At the same time they found that marketers were having to pay orders of magnitude more to buy these targeted ads, despite the minuscule additional revenue they generated for the publisher.").

81 Veronica Marotta, Vibhanshu Abhishek & Alessandro Acquisti, *Online Tracking and Publishers' Revenues: An Empirical Analysis* (May 2019) (preliminary draft), https://weis2019.econinfosec.org/wp-content/uploads/sites/6/2019/05/WEIS_2019_paper_38.pdf.

82 *See* Danielle Keats Citron, *Reservoirs of Danger: The Evolution of Public and Private Law at the Dawn of the Information Age*, 80 S. Cal. L. Rev. 241 (2006).

83 Louise Matsakis, *Online Ad Targeting Does Work—As Long As It's Not Creepy*, Wired (May 11, 2018), https://www.wired.com/story/online-ad-targeting-does-work-as-long-as-its-not-creepy.

84 Steph Mitesser, *You Can't Just Tell Everyone to Leave Facebook*, Outline (Apr. 3, 2018), https://theoutline.com/post/4040/you-cant-just-tell-everyone-to-leave-facebook.

85 *See, e.g.* Steven Melendez & Alex Pasternack, *Here Are the Data Brokers Quietly Buying and Selling Your Personal Information*, Fast Company (Mar. 2, 2019), https://www.fastcompany.com/90310803/here-are-the-data-brokers-quietly-buying-and-selling-your-personal-information.

86 Fed. Trade Comm'n, *FTC Approves Final Consent Orders Settling Endorsement and Deceptive Native Advertising Charges against Creaxion Corporation and Inside Publications, LLC* (Feb. 8, 2019),

87 47 U.S.C. § 230(c).

88 *Section 230 of the Communications Decency Act*, Elec. Frontier Found., https://www.eff.org/issues/cda230.

89 *Id.*

90 47 U.S.C. § 230(c)(2)(A).

91 *See, e.g.*, Olivier Sylvain, *Discriminatory Designs on User Data*, Knight First Amend. Inst. (2018), https://knightcolumbia.org/sites/default/files/content/Sylvain_Emerging_Threats.pdf; Rebecca Tushnet, *Power Without Responsibility: Intermediaries and the First Amendment*, 76 Geo. Wash. L. Rev. 101, 117-19 (2008).

92 AirBnB v. San Francisco, 2016 WL 6599821, 6 (N.D. Cal. Nov. 8, 2016) (upholding an ordinance that prohibited unlicensed listings on accommodation booking website because restricting the listings did not require the court to treat the website as a publisher).

93 *See, e.g.*, HomeAway.com, v. City of Santa Monica, No. 18-55367 (9th Cir. 2019) (dismissing lawsuit for failure to state a claim, but rejecting 230 immunity when a platform is required to screen for illegal content and holding that a duty to screen illegal content does not require "the Platforms to review the content" provided by third parties when the content reviewed is "distinct, internal, and nonpublic"); *accord* Fair Housing Council v. Roommates.com, LLC, 521 F.3d 1157 (9th Cir. 2008) (a website that acts as the developer of illegal selection choices may be liable for illegal conduct following those choices despite CDA 230); Brief for Upturn as Amicus Curiae supporting Plaintiffs, Onuoha v. Facebook, Inc., No. 5:16-cv-06440-EJD (N.D. Cal. filed Nov. 3, 2016) (because Facebook creates targeting categories that match protected categories, it should be liable for discrimination under the Fair Housing Act).

94 Sylvain, *supra* note 91; Danielle Keats Citron & Benjamin Wittes, *The Problem Isn't Just Backpage: Revising Section 230 Immunity*, 2 Geo. L. Tech. Rev. 453 (2018).

95 See Herrick v. Grindr LLC, 2019 WL 1384092 (2d Cir. March 27, 2019) (granting immunity under Section 230 and denying plaintiff's claims that a platform must warn a user of harmful content directed at the user).

96 Josh Constine, *Zuckerberg Implies Facebook Is a Media Company, Just "Not A Traditional Media Company"*, TechCrunch (Dec. 21, 2016), https://techcrunch.com/2016/12/21/fbonc ("It's not a traditional media company. You know, we build technology and we feel responsible for how it's used.").

JEFF GARY is an Institute Association at the Institute for Technology Law and Policy at Georgetown University Law Center.

ASHKAN SOLTANI is an independent researcher and technologist specializing in privacy, security, and behavioral economics.

FILED UNDER **ESSAYS AND SCHOLARSHIP**